Model Specification Effects in ETS/Nutrition Research

S. J. Kilpatrick

Summary

In Hirayama's study the average annual death rate for wives aged 60–69 from lung cancer is 18 per 100,000 as compared with 39 for all Japanese women (Exhibit 9). Also, wives aged 50–59 have the same lung cancer death rate as wives aged 60–69 (i.e., no age trend). These results may arise from the 23% of the cohort which is missing.

These anomalies are obscured in Hirayama (1984) by the use, in all but one table (Table 2), of husband's age rather than the wife's age.

Using wife's age to analyze wife's mortality leads to an additive model for lung cancer. The use of the relative risk is thus contra-indicated.

The weak association of husband's smoking status with wife's lung cancer mortality is probably a consequence of incomplete age adjustment when coarse age groups of 10 years are used over a 16-year period. Suggestions are made for further analyses using ungrouped information.

The effect of daily intake of green and yellow vegetables on lung cancer is also reanalyzed. A standard analysis of these data leads to different results than those given by Hirayama (1984).

Public examination of these data is called for to yield independent answers to the questions raised here.

Introduction

Hirayama (1984) reports on a longitudinal record linkage study of married women who, in 1965, were reported to be non-smokers. Interviews using form 1 (see Exhibit 1) were carried out October through December 1965 of persons 40 years and above in 49 districts in 29 health center districts in Japan. In 1971, a 3% sample of those subjects were re-interviewed (Hirayama 1982) using form 2 (Exhibit 2). Form 2 is form 1 with additional questions on current health status and illnesses in the past five years. A second follow up was apparently done between 1971 and 1983 since Hirayama (1984) refers to a recent study of 410 males and 158 females in Aichi province. Apart from these, no monitoring of the population was carried out apart from linking deaths in the period 1966 to 1981 to the original questionnaire (Exhibit 1).

The cause of death in those women who had died by 1981 was linked to the initial interview in 1965 of both husband and wife. In the sequel it is important to note that date of birth, age of first marriage, age started smoking and date of death are recorded. Linkage of a married couple's original responses to the wife's death certificate can therefore yield the woman's precise age at entry. Likewise, for a non-smoking wife, the

H. Kasuga (Ed.) Indoor Air Quality

Exhibit 1

Form 1 **Initial survey**

Health Questionnaire Name of Prefecture Health Center

District code	Household code		Individual code
Name	M F	Date of birth (year month day) 1. Single 2. Married	3. Divorced 4. Widowed
Address			
Place of birth	Prefecture	City	Occupation (in detail)
For women	Number of children	Length of breast feeding after last delivery month(s)	Age at first marriage

Anamnesis		
Eating Habits	Rice/Wheat	Amount/day Frequency
	Meat	1. Daily 2. Occas 3. Rare 4. None 5. Obscure
	Fish and shell fish	1. Daily 2 Occas 3. Rare 4. None 5. Obscure
	Milk and goat milk	1. Daily(amount) 2. Occas 3. Rare 4. None 5. Obscure
	Green-yellow vegetables	1. Daily 2. Occas 3. Rare 4. None 5. Obscure
	Pickles	1. Every meal 2. Daily 3. Occas 4. Rare 5. None 6. Obscure
	Soybean paste soup	1. Daily 2. Occas 3. Rare 4. None 5. Obscure
Favorites	Smoking	1. Smoking daily (a) Cigarette No./day (b) Kizami (c) Others 2. Occas 3. Ex. 4. None 5. Obscure Age started ( )
	Alcohol	1. Daily 2. Occas 3. Rare 4. None 5. Obscure Type (1) Sake (2) Shochu (3) Beer (4) Whisky (5) Others (6) Obscure
	Green tea	1. Very hot 2. Moderate 3. None 4. Obscure

age at which the husband started smoking and the date of the marriage can yield the duration of exposure to husband's cigarette smoke at the first interview.

The data from this study as presented by Hirayama (1984) has been summarized in 10 tables for non-smoking wives. Exhibit 3 lists these tables and shows by table number the cause of death and the factors by which the cause of specific death rates are classified. The levels of a given factor are given in parentheses. Note that Table 5 is a collapsed form of Table 6 or of Table 10 and Table 7(1) of Table 7(2). Note that Table 9 is Table 8 omitting non-smoking husbands. Only one Table, Table 2, gives wife's age group. The relationship of wife's age group to her daily intake of green/yellow vegetables is not given, nor of wife's age group to husband's age group, husband's drinking habit or husband's occupational group.

Poisson Regression

The following gives the standard analysis of the tables published in Hirayama (1984). Since Tables 5, 7(1) and 9 are all collapsed versions of other Tables, they are omitted from

Exhibit 2

Form 2 **Second survey**

Health Questionnaire Name of Prefecture Health Center

District code	Household code		Individual code
Name	M / F	Date of birth (year month day) 1. Single 2. Married 3. Divorced 4. Widowed	
Address			
Place of birth	Prefecture City	Occupation (in detail)	
For women	Number of children	Length of breast feeding after last delivery month(s)	Age at first marriage

Anamnesis		
Eating Habits	Rice/Wheat	Amount/day Frequency
	Meat	1. Daily 2. Occas 3. Rare 4. None 5. Obscure
	Fish and shell fish	1. Daily 2. Occas 3. Rare 4. None 5. Obscure
	Milk and goat milk	1. Daily(amount) 2. Occas 3. Rare 4. None 5. Obscure
	Green-yellow vegetables	1. Daily 2. Occas 3. Rare 4. None 5. Obscure
	Pickles	1. Every meal 2. Daily 3. Occas 4. Rare 5. None 6. Obscure
	Soybean paste soup	1. Daily 2. Occas 3. Rare 4. None 5. Obscure
Favorites	Smoking	1. Smoking daily (a) Cigarette No./day (b) Kizami (c) Others 2. Occas 3. Ex. 4. None 5. Obscure Age started ( )
	Alcohol	1. Daily 2 Occas 3. Rare 4. None 5. Obscure Type (1) Sake (2) Shochu (3) Beer (4) Whisky (5) Others (6) Obscure
	Green tea	1. Very hot 2. Moderate 3. None 4. Obscure Others (1. Tea 2. Coffee 3. Cola 4. Cider)
Current Health Status (danger signals)		1. Stomach trouble, indigestion, no appetite, change in food choice. 2. Vaginal discharge, irregular bleeding. 3. Lump in the breast 4. Difficulty in swallowing. 5. Blood or mucos in stool. 6. Continued cough, bloody sputum, hoarseness. 7. Chronic ulcer in the mouth/skin. 8. Difficulty in urination, blood in urin. 9. Irritation/uneasiness 10. Difficulty in sleeping. 11. Heart trouble.
Currently		1. Healthy 2 In bed (by) from when.
Major illness during past 5 years		name of illness time duration. 1) 2)
Health Check		1 none 2 yes (stomach X ray chest X ray blood pression others)

analysis. Note that, because of different groupings of husband's occupational group, Table 3 cannot be derived from Table 8, nor Table 6 from Table 10. Indeed, since person years are not given, the study appears to call for a Proportional Mortality Analysis of lung cancer, other cancer and ischemic heart disease mortality in non-smoking wives, cross-classified by wife's age group (4) × husband's age group (4) × husband's smoking classification (5) × husband's drinking habit (4) × husband's occupational group (10) ×

Exhibit 3

Tables as presented in Hirayama(1984)

TABLE	OUTCOME	FACTORS*
1	LCD	HAGE(4) x HCIG(5)
2	LCD[1]	WAGE(4) x HCIG(3)
3	LCD	HAGE(4) x HCIG(3) x HOCC(10)
4	LCD	HAGE(4) x HALC(4)
5	IHD[2]	HAGE(4) x HCIG(3)
6	IHD	HAGE(4) x HCIG(3) x HOCC(10)
7(1)	OTHCA[3]	HAGE(4) x HCIG(3)
7(2)	OTHCA	HAGE(4) x HCIG(3) x HOCC(10)
8	LCD	HAGE(4) x HCIG(3) x HOCC(2) x GYV(2)
9	LCD	HAGE(4) x HCIG(2) x HOCC(2) x GYV(2)
10	IHD	HAGE(4) x HCIG(3) x HOCC(2) x GYV(2)

1 LCD.......lung cancer deaths

2 heart disease deaths

3 cancer deaths

*Factors:

HAGE husband's age group

WAGE wife's age group

HCIG husband's daily smoking habit

HALC husband's daily alcohol intake

HOCC husband's occupational group

GYV wife's daily intake of green & yellow vegetables

Levels: In a factor XXX(n), n is the number of levels of factor XXX in the specified table

wife's daily intake of green/yellow vegetables (2). Parenthetically there is no reason today why, with modern computing techniques this basic tabulation should not have been analysed directly, instead of piecemeal as reported. Unfortunately the basic data is not available to the author (Hirayama, personal communication).

The analysis which follows is that recommended by Breslow & Day (1986) for cohort studies. In the absence of person years, the cumulative mortality rate over the period 1966–1981 is used as the response variate. (This assumes no "competing" causes of death and no loss to follow-up. This rate is not strictly a risk estimate since it depends on the duration of the study, the period of the study and on the choice of study population). A Poisson error structure is specified with a logarithmic link function which is the default for a Poisson error structure in GLIM (Payne 1985). The regressions are weighted according to the number of non-smoking wives in each cell.

Following Breslow (1987), nuisance variables, irrespective of their technical significance, are fitted before the factor of interest, i.e. either husband's smoking status or green and yellow vegetable intake. A factor is considered to have a significant association with the specified mortality rate only if the deviance reduction in the model, for the degrees of freedom associated with fitting that factor, is significant at the 5% level. (Note that Dr. Hirayama here uses a one-sided 10% level of significance which is equivalent to a two sided test at 20% significance). Analysis of residuals and regression diagnostics are not given here. Rather the model fit is evaluated using the approximation of the residual deviance and its degrees of freedom to the χ^2 distribution which "may overstate the degree of departure from the fitted model when many cells contain small counts" (Breslow & Day 1986, p. 137).

Husband's Age Group, Husband's Drinking Habit and Lung Cancer Mortality

Age at interview is clearly a powerful factor which must be fitted first. In some tables age exhibits a powerful linear trend and can be fitted as a single numeric variable. Where this is possible it is done to achieve the most parsimonious model.

Table 4 gives a cross classification of husband's age group × husband's drinking habit for lung cancer mortality in non-smoking wives. It is surprising that other "nuisance factors" are not included. Nevertheless, this table is analysed first, largely to investigate the association of lung cancer mortality with husband's age group.

Standard Poisson regression of Table 4 confirms that husband's age group is an important factor in lung cancer mortality (see Exhibit 4). Husband's age group exhibits a strong linear trend with lung cancer mortality. A log-log plot of lung cancer mortality rate vs husband's age group however gives a slope less than 3 whereas a slope of 4 has been reported for non-smokers using attained age (Seidman 1985). Husband's drinking habit shows no significant association in this table with lung cancer mortality but no adjustment has been made for other nuisance variables. Thus one would expect an association between husband's drinking and smoking habits.

ETS and Lung Cancer Mortality (Tables 1, 2, 3)

The only measure of ETS exposure given is husband's smoking classification, the number of cigarettes reported in 1965 as smoked daily by the husband. The standard practice of demonstrating that a factor is significant before looking for a trend is followed here. As for husband's age group in Table 4 above, husband's smoking classification shows a strong linear trend in certain tables and is entered as a single numeric variable in the interests of parsimony where possible.

"Typical practice is to consider 5 year intervals of age and time so as to be able to study variation in rates" (Breslow & Day 1980, p. 47–48). Hirayama (1984) uses 10 year age groups and does not divide the 16 year period. In general, an age classification of 10 years at entry in a study lasting 16 years with no time dependent factors may mean that the age effect has been incompletely adjusted (Mantel 1983). Thus, for a lung cancer mortality rate which rises exponentially with age, it is plausible that the significance of husband's smoking classification is an indication of incomplete age adjustment, given the rapidly changing habits of cigarette smoking in the period before 1966 (Kristen 1986).

Note also that duration of ETS exposure is confounded to an unknown extent with age at first interview. Thus, in the absence of other information, assume a constant age at

Exhibit 4

Summary of the fit of the best† multiplicative model

OUTCOME TABLE	PREDICTIVE MODEL	DEVIANCE(d.f.)
LCD 1	A	13.0 (18)
	A+HCIG	3.7 (14)
LCD 2	WAGE	16.8 (8)
	WAGE+HCIG	10.7 (6)
LCD 3	A+HOCC+C	71.9 (108)
LCD 4	A	15.3 (14)
IHD 6	HAGE+HOCC	115.1 (106)
	HAGE+HOCC+HCIG	109.6 (104)
OTHCA 7(2)	HAGE+HOCC	134.7 (106)
LCD 8	A+HOCC+C	50.8 (44)
IHD 10	HAGE+HOCC+HCIG	50.7 (41)

Key:

A is husband's age fitted as a linear trend

C is husband's daily smoking habit fitted as a linear trend

other factors, outcomes as defined in Exhibit 3

†Best in the sense of minimum residual deviance after fitting all 'nuisance' parameters as factors or (if warranted) as trends and then fitting the explanatory variables, HCIG [Tables 1, 2, 3, 4, 7(1)] or GYV [Tables 8, 10].

marriage and at starting smoking. In 1965, older non-smoking wives of smoking husbands will have been exposed to ETS for a longer period than younger wives. (This expectation of an increased relative risk for older wives is not evident from an analysis of Table 2 (see Exhibit 7)). Form 1 (Exhibit 1) records the age at which the husband started smoking. Given this and the date of the marriage from a linked wedding certificate, it should be possible to estimate the duration of ETS exposure by the non-smoking wife of a smoking husband prior to 1966 as well as the wife's age at first exposure.

Table 1, like Table 4, gives the cross-classification of husband's age group × husband's smoking status. As shown in Exhibit 4, only husband's age group is significant and exhibits a strong trend, as in Table 4. Although husband's smoking classification is not significant, it approaches significance ($\chi^2 = 9.3$ on 4 degrees of freedom, P just greater than 5%).

The effect of re-classifying husband's smoking classification from 5 levels to 3 levels can be seen in Table 3 which also gives a breakdown by 10 occupational groups, HOCC (10). Although husband's occupational group with 9 degrees of freedom is not significant, husband's smoking classification with 3 levels now is. Indeed husband's smoking classification with three levels now exhibits a strong trend (Exhibit 4).

Table 2 is unique in this publication in that lung cancer mortality is adjusted for wife's age group. Indeed this appears to be the only occasion on which Hirayama has included wife's age group in an analysis in any of his many publications from this study. (We shall see that husband's age group is not a surrogate for wife's age group.)

Standard Poisson regression of Table 2 as presented shows (Exhibit 4) that wife's age group, while a significant factor, does not exhibit a trend against lung cancer mortality. Again husband's smoking classification is on the borderline of statistical significance as judged by the change in the deviance ($\chi^2 = 6.1$ on 2 degrees of freedom). Clearly the evidence for a significant relationship between lung cancer mortality and husband's smoking classification is ambivalent even without considering the influence of non-sampling errors and confounding factors.

ETS and Ischemic Heart Disease (Tables 5, 6)

The consideration of multiple outcomes for associations with ETS indicates the multivariate nature of the analysis and the lack of prior hypotheses in this study. One should allow for multiple or repeated tests of significance in the evaluation of these results.

Table 6 gives a tabulation of ischemic heart disease mortality by husband's age group, husband's smoking classification and husband's occupational group. After adjustments for both husband's age group and husband's occupational group are made (Exhibit 4), husband's smoking classification is just non-significant by the established criteria ($\chi^2 = 5.6$ on 2 degrees of freedom). This is in contrast to Table 5 (not shown) which is Table 6 collapsed over husband's occupational group, showing some confounding between husband's occupational group and husband's smoking classification for ischemic heart disease.

ETS and Other Cancers

Table 7(2) classifies other cancer against husband's age group, husband's smoking classification and husband's occupational group, the same classification as for lung cancer mortality (Table 3) and for ischemic heart disease (Table 6). This again points out that a Proportional Mortality Analysis is the preferred method of analysis here. A univariate log linear analysis confirms that husband's occupational group is significantly associated with other cancer mortality. This association is almost entirely due to husband's occupational group 5, "farmers, laborers and fishermen" which has an estimated relative risk of 1.45 with 95% confidence limits of (1.04–2.03). No significant association with husband's smoking classification is detected with other cancer.

Green/Yellow Vegetables and Lung Cancer Mortality (Table 8)

Switching the focus now from husband's smoking classification to daily intake of green/yellow vegetables, we first fit all factors other than daily intake of green/yellow vegetables in Table 8 as nuisance parameters. The analysis of deviance reduction establishes that daily intake of green/yellow vegetables has a non-significant association (Exhibit 4).

Green/Yellow Vegetables and Ischemic Heart Disease (Table 10)

Table 10 gives ischemic heart disease mortality by husband's age group, husband's smoking classification, husband's occupational group and daily intake of green/yellow vegetables.

No significant association is found (Exhibit 4) with daily intake of green/yellow vegetables, after adjustment for these other factors ($\chi^2 = 0.6$ on 1 degrees of freedom).

In summary, standard Poisson regression, using the conventional 5% level of significance indicate, on the basis of these published tables,

- husband's smoking classification is marginally associated with wife's lung cancer mortality, the size of the effect being of borderline significance and dependent on the presence or absence of other factors in the model and the number and grouping of classes used in the husband's smoking factor.
- that husband's drinking habit shows no significant association with lung cancer mortality in the limited data published here.
- that daily intake of green/yellow vegetables shows no significant association with lung cancer mortality or with ischemic heart disease mortality.
- that husband's smoking classification is of borderline significance with wife's ischemic heart disease mortality.
- that husband's smoking classification shows no significant association with other cancer mortality.

These findings may be compared against those of the original report. There Hirayama (1984) claims "a significantly increased risk of" lung cancer mortality "in relation to the extent of the husband's smoking ... The association was significant when observed by age of husbands ... and also by age of wives." "Similar significant risk elevation of lung cancer with the increase in the extent of husband's smoking was observed with ischemic heart disease when observed by husband's age group and husband's occupational group."

"The risk-reducing effect of daily intake of green-yellow vegetables on lung cancer was observed for passive smoking ... Those women eating green-yellow vegetables daily showed a significantly lower risk of lung cancer from the passive influence of their husbands' smoking."

Power Fit

Exhibit 4 which summarizes the best fitting multiplicative model indicates that in some instances this fit may not be too good (or that interaction terms are necessary). Thus, the residual deviance considered as an approximate χ^2 indicates that for both models fitted toTable 2, the fit is of borderline significance. This is true also of Table 7 (2), Table 8 and

Exhibit 5

Deviances (df) for additive, multiplicative and best fitting power models

OUTCOME TABLE	ADDITIVE	POWER	MULTIPLICATIVE
LCD Table 1	3.79 (12)	2.76 (12)	3.24 (12)
LCD Table 2	7.91 (6)	7.85 (6)	10.72 (6)
LCD Table 3	2.57 (6)	1.50 (6)	1.95 (6)
IHD Table 5	2.32 (6)	2.17 (6)	2.72 (6)
OTHER CA Table 7(1)	3.28 (6)	2.46 (6)	3.68 (6)

Model is 1+HAGE(4)+HCIG(5) for Table 1
.... 1+WAGE(4)+HCIG(3) for Table 2 and
.... 1+HAGE(4)+HCIG(3) for Tables 3, 5 and 7(1)

Rates are DTHS/POP (1966-1981)
for LCD (Tables 1,2,3), OTHER CA (Table 7(1)), IHD (Table 5)

Table 3 has been collapsed over HOCC

Table 10. This test is approximate. Nevertheless, it was decided to investigate the best fitting power model (Breslow 1986) to these tables. The goodness of fit of the additive, multiplicative and best fitting power model to these data are compared in Exhibit 5 in terms of residual deviance. Note that the additive and multiplicative models are special cases of the power model with exponents equivalent to one and zero respectively.

In Exhibit 5, an attempt has been made to fit the same predictive equation, adjusting for age and ETS exposure across the different sets given in Hirayama (1984). Overall, husband's age in 1965 classified by 10 year age groups, gives very satisfactory fits, irrespective of which Poisson model is used. In contrast, Exhibit 5 shows that poor fits result from the use of wife's age in 1965, classified in 10 year age groups, the multiplicative model giving the worst fit.

Exhibit 5 also reveals that the power-deviance curve is generally quite flat. Apart from Table 2, for which the additive model is the model of choice, the data, as presented in Hirayama (1984), do not discriminate well between additive and multiplicative models.

Exhibit 6

POWER VALUE FOR BEST MODEL

POWER

OUTCOME Table	ρ	$\hat{\rho}$	ρ
LCD Table 1	1	0.40	0
LCD Table 2	1	1.14	0
LCD Table 3	1	0.39	0
IHD Table 5	1	0.61	0
OTHER CA Table 7(1)	1	-	0

predictive equations as in Exhibit 5

- $\hat{\rho}$ meaningless since HCIG has zero estimates

The power value in the best fitting model is given in Exhibit 6. This lies between $\rho = 1$, the additive model and $\rho = 0$ (which is equivalent to the multiplicative model) for all but Table 2. The best fitting power model for Table 2 is larger than 1.00, indicating that the multiplicative model as fitted above, is contra-indicated. An additive model, then, is clearly preferred over the multiplicative model for Table 2, which, alone, uses wife's age group. However, the flatness of the deviance curve against ρ may indicate that the assumption of a Poisson error term is incorrect.

This finding may be interpreted in biological terms and in terms of information content. Although this study is considered to be one of the largest on ETS and lung cancer and contributes heavily to any meta-analysis estimate of passive-smoking effect (NRC report. 1986) it contains little information because of the absence of specific exposure, person year and time dependent data.

Wife's Age (Table 2)

Having shown that the additive model is the model of choice for Table 2, we now consider this analysis more fully. Unfortunately we are restricted to this one simple cross classification of wife's age group by husband's smoking classification using coarse intervals and omitting others factors. Under the additive model, wife's age group and husband's smoking classification are significant factors ($\chi^2 = 32.6$ on 3 degrees of

Exhibit 7

Table 2 Lung Cancer Relative Risk (1966–1981)

HUSBAND'S SMOKING

		Non	Ex or 1 – 19/d	20 + /d
WIFE'S AGE	40–49	1.0	2.4	3.3
	50–59	1.0	1.6	1.9
	60–69	1.0	1.2	1.0
	70–79	1.0	0.1	0.5

freedom for wife's age group and 8.84 on 2 degrees of freedom for husband's smoking classification, $0.05 > P > 0.01$).

Although this is an improvement over the multiplicative model, the residual deviance is 7.91 on 6 degrees of freedom, indicating that this model may still not be good fit. Likewise the best fitting power model had residual deviance of 7.85 on 6 degrees of freedom – not a great improvement.

This is paradoxical. As we move from husband's age group to wife's age group (which should give a more direct relationship between age and lung cancer mortality) we, in fact, find continued evidence of an interaction between wife's age group and husband's smoking classification, irrespective of which model we use. It must be concluded that Table 2 contains insufficient detail in wife's age group and exposure to ETS or that other factors, not shown, are associated with lung cancer mortality in the non-smoker.

An alternative way of explaining why the multiplicative model is not the model of choice when wife's age is used is to examine the relative risks. The use of the multiplicative model assumes that the relative risk is constant with age. Exhibit 7 however demonstrates a clear trend in the relative risk which falls from values above 1 at young ages to values below 1 over 70. These trends arise because of the different effects of age in the three smoking status categories. Clearly (as may be seen from Exhibit 8 (figure)) the rates for the three smoking status categories are approximately equal at wife's age 60–69 but differ (in different directions) at other ages. Exhibit 9 compares average annual rate by wife's age (Table 2) with the same rate when classified by husband's age (Table 1) and both are compared with estimated Japanese rates for females. The rates for Table 1 and Table 2 are both uniformly lower than Japanese rates for women. Either wives have a much more favorable experience than all women or Hirayama's study subjects are unrepresentative of Japanese wives or both. In addition Exhibit 9 reveals an anomaly in Table 2 in that, unlike Japan or Table 1, the lung cancer death rates when classified by wife's age show no age trend from age group 50–59 to 60–69!

This suggests a serious misclassification of wife's age, wives who were 50–59 being recorded as 60–69 at the initial interview. Alternatively, and more likely, lung cancer deaths for wives aged 60–69 at initial interview are seriously under reported, giving a spuriously low average annual year lung cancer death rate of 18 per 100,000 as compared with a Japanese rate for all women of 39.

Turning now to an examination of the selected cohort, we look at the percentage distribution of husband's smoking status by wife's age. Exhibit 10 gives a 1965 cross-

2023513441

Exhibit 8

Table 2 Wife's Lung Cancer Death Rate (logarithm) (1966-1981) by Husband's Smoking Status (1965)

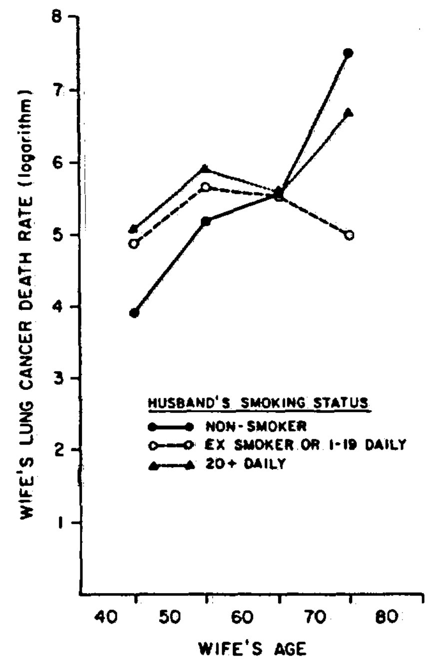

Exhibit 9

COMPARISON OF RATES BY AGE

Annual Average Lung Cancer Death Rate /100,000
1966-1981

AGE	JAPAN†	TABLE 1*	TABLE 2
40-49	12	7	8
50-59	20	12	18
60-69	39	23	18
70-79	46	36	35

† based on Japanese rates for women. (Segi et al., 1981)

* husband's age.

2023513442

Exhibit 10

Table 2 Distribution of Husband's smoking status by Wife's age

HUSBAND'S SMOKING

		Non	Ex or 1 – 19/d	20 + /d	Total
	40–49	21%	46%	33%	38,025(100%)
WIFE'S	50–59	24%	49%	27%	32,089(100%)
AGE	60–69	30%	51%	19%	20,344(100%)
	70–79	16%	62%	22%	1,082(100%)

Exhibit 11

Table 1 Distribution of husband's smoking status by his age

HUSBAND'S SMOKING

		Non	Ex	1 – 14/d	15 – 19/d	20 + /d	TOTAL
	40–49	19%	4%	27%	16%	34%	32,027(100%)
HIS	50–59	23%	6%	29%	12%	30%	33,253(100%)
AGE	60–69	29%	11%	30%	10%	19%	24,214(100%)
	70–79	37%	17%	30%	5%	11%	2,046(100%)

sectional view of cohort changes in husband's smoking habits. A number of points arise. Although Dr. Hirayama has published no information from his 3% re-interview survey on changes in smoking status between 1965 and 1971, such changes in smoking status occurred and may be of the order demonstrated in Exhibit 10 for husbands. (We have no information on wife's changes in smoking habits. Dr. Hirayama claims that 1.96% of the women polled in his 3% re-interview survey were misclassified as to smoking status. It is difficult to understand how he can discriminate between conversion from non-smoking to smoking status given the nature of the smoking question revealed in Exhibits 1 and 2. If 1.96% of wives were misclassified, what is the conversion rate from non-smokers to smokers in the period 1965–1971 among these wives?)

Secondly, the intermediate smoking classification (Ex or 1–19/d) is the most numerous of the three smoking status classifications for the husband and is a composite of ex-smokers and light and intermediate smokers (1–14/d and 15–19/d). It could be argued that as the most numerous the intermediate group should be used as the baseline for testing the significance above and below these rates for non-smokers and heavy smokers (20+/d) respectively. However, this group of wives has an unknown mixture of exposures to passive smoking. As indicated above, form 1 (see Exhibit 1) records

information on duration of exposure of a wife to her husband's smoking but this has never been used in Dr. Hirayama's many publications.

Finally, a husband's smoking status is clearly dependent on his age (Exhibit 11). Thus, as a surviving husband ages he is less likely to be classified as a 20+/d smoker and most likely to be classified as a non-smoker or ex-smoker. Dr. Hirayama groups ex-smokers with light smokers (what happens to "occasional smokers"? (see form 1 (Exhibit 1)). In terms of exposure before 1966 this is correct but it may be argued that ex-smokers should be grouped with non-smokers since wife's exposure is zero after 1965 and lung cancer latency is of the order of 10 years.

Better still, fine detail should be preserved in order to allow for the true expression of factors and covariates. Thus it is likely that the association of husband's smoking status with wife's lung cancer mortality is simply an example of incomplete age adjustment, using 10 year age groups with a 16 year cumulative mortality. In other words, husband's smoking status is confounded with wife's age. Again this can be remedied by using modern analytical techniques to analyse the data in detail.

Discussion

Dr. Hirayama's publications, over the years, have analyzed this longitudinal record linkage study from many aspects. Given the nature of the study and the absence of specific details, it is clear that these data can not be used to confirm hypotheses or to strengthen the evidence for or against a causal mechanism between causes of death (his outcomes) and his factors, since "we can be easily misled by variables not represented or recognized in a study" (Tukey & Mosteller 1977, p. 119) and since "tests of significance and confidence intervals that fail to account for the lack of fit of a given model may be seriously misleading " (Breslow 1987, p. 37).

The absence of relevant factors and specific details is shown here in the inability of these published data to discriminate between additive and multiplicative Poisson regression models. It is unfortunate that the Committee on Passive Smoking (NRC 1986) gave so much weight to Dr. Hirayama's conclusions in their review of the evidence for and against passive smoking as a cause of lung cancer.

This standard re-analysis of Hirayama (1984) points to husband's smoking status being a surrogate for some other factor or factors. Thus, an unadjusted analysis of husband's alcohol intake showed no association with lung cancer mortality. If husband's smoking status were a causal factor in the formation of lung cancer, one would expect alcohol intake also to be associated with this risk because of the association of smoking and drinking habits.

Comparison with other cohort studies shows how approximate the evaluation of ETS exposure is in this study. Thus, for example, Smith & Doll (1982), investigating the effect of irradiation on leukemia mortality use both age at first exposure and duration since first exposure as factors. Dr. Hirayama has linked his initial interview file with death certificates for selected causes of death. It should be possible to link wedding, divorce and death certificates (for all causes) to the original file in order to estimate the duration of the marriage. Further, since the age at which the husband started smoking was recorded, the duration of the wife's exposure to passive smoking could be estimated. This assumes that no non-smoking wife started smoking in the interval 1966–1981. Figure 1 of Hirayama (1984) and Kristen (1986) show a rapid rise in per capita cigarette consumption in this period in Japan. In the light of this increase, it is plausible to assume that a number of these wives became smokers after 1965. More non-smoking wives of smoking husbands

2023513444

would be expected to become smokers than among those married to non-smokers because of the husband's example. Likewise, more wives of smokers are likely to have been misclassified as non-smokers in 1965 than wives of non-smokers (Lee, in press).

In the absence of information on duration of exposure, we know only the reported smoking status of husband and wife at initial interview. Assuming stability throughout, older wives in 1965 have been exposed for longer than younger wives. If so, relative risks should increase but the opposite is true (Exhibit 7). Indeed, as has been indicated, wife's age in 1965, is a less effective explanatory variable for cumulative lung cancer mortality than husband's age. Dr. Hirayama's analyses which use the spouse's age for age standardization are of questionable value. Theories of carcinogenesis relate the incidence of cancer to the age of the experimental animal or the individual. The analytical comparisons given here indicate that husband's age *cannot* be used as a surrogate for wife's age if the age at entry of the decedent is used. The importance of this conclusion may be seen in the observation that, if Dr. Hirayama's study is excluded from a global estimate of passive smoking effects on lung cancer, the resultant meta-analysis gives a value which is not significantly greater than 1!

Dr. Hirayama's study ascertained 142,857 women 40 years or over in 1965. Figures 7 and 8 of Hirayama (1984) document the smoking history or exposure of 108,906 females, leaving 33,951 women unaccounted for. It might be assumed that 24% of the female cohort were widows in 1965 except for Dr. Hirayama's statement "information on the smoking history of the husbands of non-smoking women with lung cancer was available – in 77.3% of cases (174 out of 240)" (Hirayama 1981). This means that the 91,540 wives analysed here and in Hirayama (1984) represent 77.3% of a total of 118,422 wives in 1965. Clearly it is impossible to re-construct the total female cohort from the information given. If, as stated by Dr. Hirayama, 23% of his study group are missing, then his confidence limits are too narrow in that they do not allow for the effect of these non-sampling errors. Inclusion of non-sampling errors for the 23% missing wives totally negate his claims of significance for the association between passive smoking and lung cancer and between green and yellow vegetable intake and lung cancer.

This investigation prompts the author to call for an international panel of scientists to be given access to Dr. Hirayama's files. An independent evaluation is needed of the contribution which this unique study can make to the role of passive smoking and dietary habits in the etiology of lung cancer and heart disease.

Acknowledgement: The author is indebted to Dr. John Viren for his suggestions, criticisms and advice.

References

Breslow N (1986) Use of the power transform to discriminate between additive and multiplicative models in epidemiologic research. In: Moolgavkar SH, Prentice Ross L (eds) Modern statistical methods in chronic disease epidemiology. Wiley, New York, pp 181–196

Breslow NE (1987) Logistic regression, proportional hazards and related methods for the analysis of chronic disease risk. Presented at the German-American Seminar on Longitudinal Studies and their Analyses, Institute for Social Medicine and Epidemiology of the Federal Health Office, Berlin. June 23–26, 1986. Revised January, 1987

Breslow NE, Day NE (1980) Statistical methods in cancer research, vol 1. The analysis of case-control studies. IARC Scientific publications Number 32. Lyon, France.

Breslow NE, Day NE (1986) Statistical methods in cancer research, vol 2. The design and analysis of cohort studies. IARC Scientific publications Number 82. Lyon, France

Hirayama T (1981) Modern medicine (interview). Münchener Medizinische Wochenschrift 123(40):1480–1488

Hirayama T (1982) Does daily intake of green-yellow vegetables reduce the risk of cancer in man (an example of the application of epidemiological methods to the identification of individuals at low risk). In: Bartsch H, Armstrong B, Davis W (eds) Proc. of the symposium on factors in human carcinogenesis IARC sc publ no 39:531–540. Lyon, WHO

Hirayama T (1984) Lung cancer in Japan: Effects of nutrition and passive smoking. In: Mizell N, Corea P (eds) Lung cancer: causes and prevention. Chemie International, pp 175–196

Kristen MM (1986) Japanese lung cancer mortality rates 1947–1980 and per capita cigarette consumption in Japan. Int J Epidemiol 15:140–141

Lee PN (in press) Increased risk of lung cancer in non-smokers married to smokers: A result of ETS exposure or of bias? Int Archives of Occ & Enviror Hlth

Mantel N (1983) Epidemiologic investigations – care in conduct, care in analysis, and care in reporting (editorial). J Ca Research Clin Oncology 105:113–116

NRC Report (1986) Environmental tobacco smoke. Measuring exposures and assessing health effects. Committee on Passive Smoking (Barbara Hulka, Chairman). National Academy Press, Washington, DC

Payne CD (1985) The GLIM systems release 3.77. Numerical Algorithms Group, Oxford

Segi M, Tominaga S, Aoki K, Fujimoto I (1981) Cancer mortality and morbidity statistics. Japan and the world. Gann monograph on Cancer Research No 26 Japan Scientific Societies Press, Tokyo

Seidman H (1985) Age at exposure versus years of exposure. In: Garfinkel L, Ochs O, Mushinski M (eds) Selection, follow-up, and analysis in prospective studies: A workshop. NCI Monograph 67, USDHSS, NCI, Bethesda, Md, pp 205–209

Smith P, Doll R (1982) Mortality among patients with ankylosing spondylitis after a single treatment course with X rays. Br Med J 284:449–460

Tukey F, Mosteller JW (1977) Data analysis and regression. Addison-Wesley, Reading